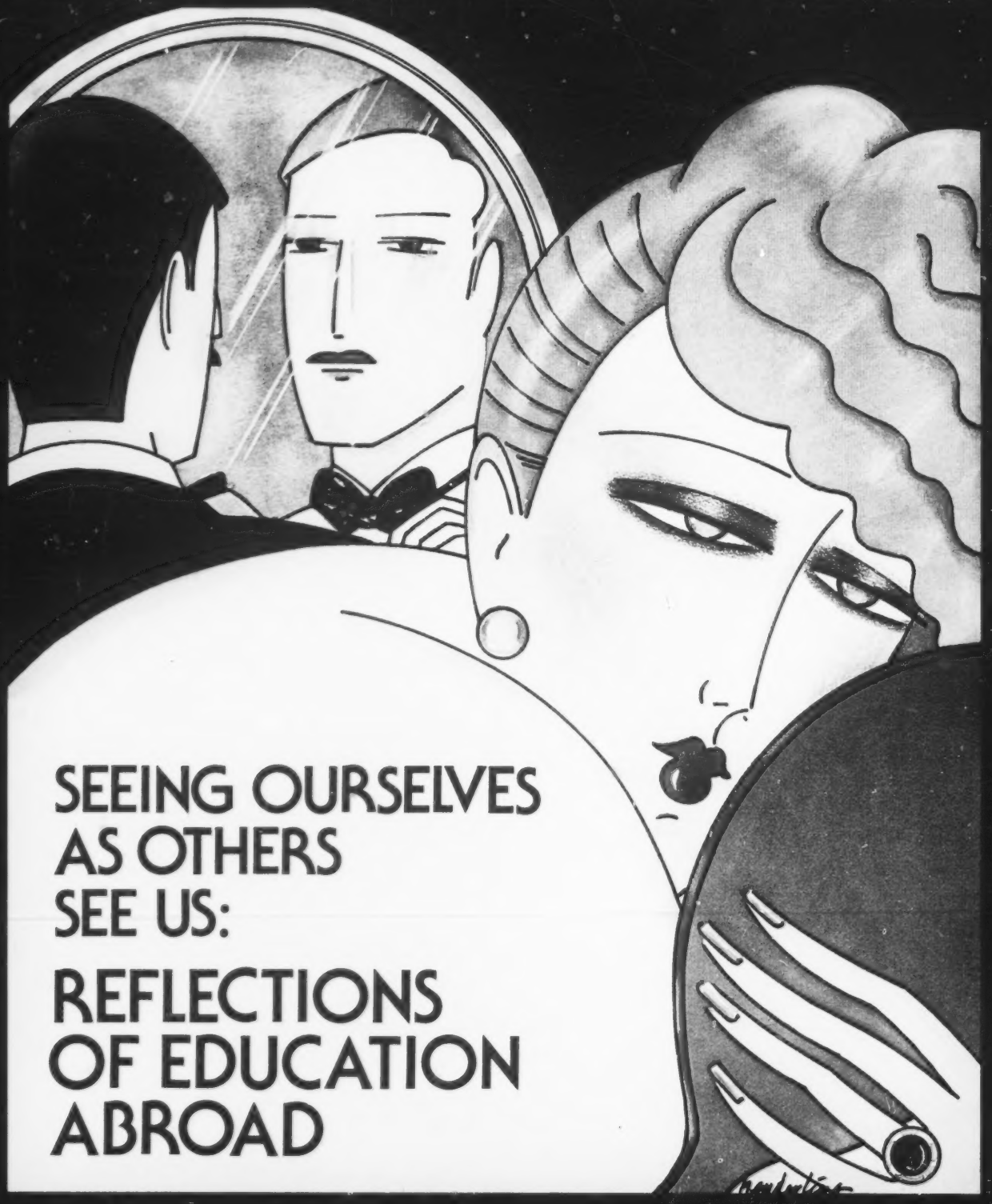
A8680-1

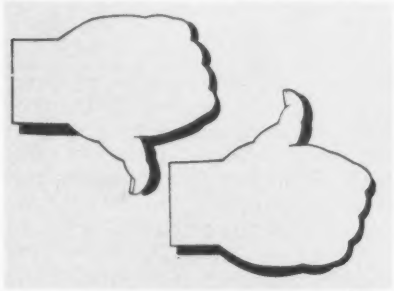
PHI DELTA KAPPAN FEBRUARY 1985

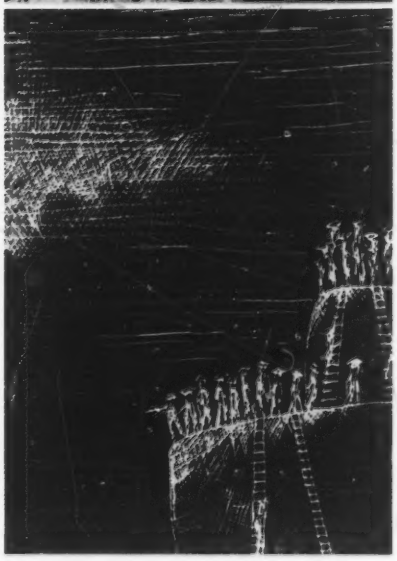
KAPPAN



SEEING OURSELVES AS OTHERS SEE US:

REFLECTIONS OF EDUCATION ABROAD





Phi Delta February 1985 • Volume 66 • Number 6

KAPPAN

FEATURES

391 **The High School and the University: What Went Wrong in America, Part 1,** *by Burton R. Clark*

398 **The School in the Achievement-Oriented Society: Crisis and Reform,** *by Torsten Husén*

403 **International Comparisons of Cognitive Achievement,** *by James S. Coleman*

407 **Mathematics Achievement in U.S. Schools: Preliminary Findings from the Second IEA Mathematics Study,** *by Kenneth J. Travers and Curtis C. McKnight*

414 **The Second International Science Study: U.S. Results,** *by Willard J. Jacobson and Rodney L. Doran*

418 **Japanese Education and Its Implications for U.S. Education,** *by Nobuo K. Shimahara*

422 **Education Reform and the Swinging Pendulum,** *by Carl Kaestle*

424 **America's Overseas School System,** *by Donovan R. Walling*

427 **Is Yours a Thumbs-Up or a Thumbs-Down School?,** *by Maryruth K. Nivens*

430 **Parent Participation in the Schools: Are the Benefits Worth the Burdens?,** *by Marilyn Johnston and Joanne Slotnik*

434 **Reporting Child Abuse and Neglect: Your Responsibility and Your Protections,** *by Bruce Beezer*

DEPARTMENTS

386 **THE EDITOR'S PAGE,** The Face in the Mirror

387 **WASHINGTON REPORT,** *by Anne C. Lewis*

389 **STATELINE,** *by Chris Pipho*

437 **PEER,** An Interview with Thomas Payzant, *by Willard Duckett*

441 **DE JURE,** Ten Years After *Goss* v. *Lopez*: An Interview with Peter D. Roos, *by Thomas J. Flygare*

443 **IN CANADA,** A Survey of the Island Province, *by Stirling McDowell*

444 **PROTOTYPES,** Project ABLE: Help for Adults with Learning Disabilities, *by William J. Vaugh*

445 **BOOKS,** *by Vernon H. Smith*

447 **NEWSNOTES**

449 **BACKTALK**

Cover Illustration: Lisa Henderling

International Comparisons Of Cognitive Achievement

by James S. Coleman

Major variations in national school systems can significantly affect students' achievement, says Mr. Coleman, who makes his point by using data from the International Association for the Evaluation of Educational Achievement.

MY FIRST SERIOUS attempt (with others) to study the effects of the schools on achievement was published in 1966 under the title, *Equality of Educational Opportunity*. That study examined the cognitive achievement at five grade levels of more than 700,000 students enrolled in schools representative of the whole breadth of U.S. elementary and secondary education.

To our great surprise, my colleagues and I found little relationship between school input and student achievement. Among all the factors that we measured, those with the largest effects on achievement had to do with students' family backgrounds. These factors included parents' educational levels and socioeconomic status, family size, and the amount of interest that parents took in their children's education and in the written word.

Another set of variables ran a poor second to family background but still exerted a stronger influence on students' achievement than school input: the family backgrounds of other students in the school. Children tended to perform better academically when they attended school with other students from educationally oriented home backgrounds. The effects on achievement of variations in the resources brought to schools by school boards, principals, and teachers were barely detectable.

At about the same time that we arrived at this conclusion, Gilbert Peaker reached a similar conclusion from an analysis that he conducted for the Plowden Report in England. Since then, many studies in the U.S. have provided evidence that variations in family background exert a stronger effect on students' achievement than do variations in school characteristics. And the results of the six-subject survey conducted by the International Association for the Evaluation of Educational Achievement (IEA) did not differ extensively; few, if any, identifiable inputs into schools could be found that brought consistently higher academic achievement.

More recent research findings indicate a somewhat different state of affairs. Stephen Heyneman and William Loxley have shown that, in less-developed countries, a reversal takes place, with school input affecting achievement appreciably and family background playing a much smaller role.[1] My own comparison of public and private high schools in the U.S. turned up achievement differences that cannot be explained by variations in students' family backgrounds or in their prior achievement. However, some indicators suggest that these achievement differences can be explained by variations in the academic demands that the two kinds of schools place on students and by variations in the schools' disciplinary standards. Andrew Greeley has pursued this finding further.[2] His data show that the academic demands placed on students by their schools strongly influence students' achievement levels.

ALL OF THIS is background to the international comparisons of cognitive achievement on which I will focus here. Such international comparisons were virtually impossible prior to the extraordinary studies conducted by the IEA. The IEA began in the Sixties with the testing of achievement in mathematics among 13-year-olds and high school seniors; it conducted a six-subject survey in the Seventies and has tested both mathematics and science achievement in the Eighties.

I intend to suggest here a way to use these IEA studies to learn more about the effects of variations in schooling on cognitive achievement. My use of the IEA data in this fashion involves a combination of three elements.

JAMES S. COLEMAN is a professor of sociology at the University of Chicago. *Another version of this article will appear in a festschrift for Professor Torsten Husén, edited by Neville Postlethwaite and published by Almqvist and Wiksell in Stockholm.*

Figure 1. Average Achievement Test Scores of 10-Year-Olds in 10 Countries — Science and Reading Comprehension

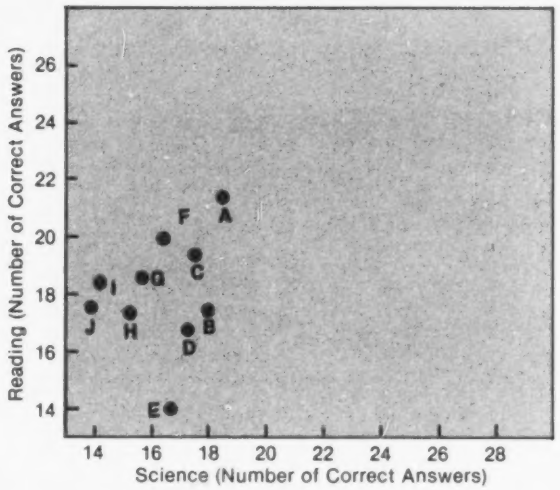
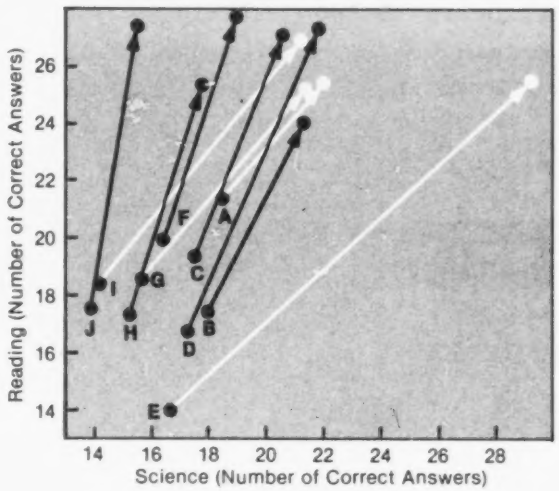


Figure 2. Average Achievement Test Scores of Both 10-Year-Olds and 14-Year-Olds in 10 Countries



• It involves looking for the effects of variations in schooling not by comparisons *within* a country (the principal way in which IEA data have been analyzed in the past) but by comparisons *between* countries.

• It involves using a younger cohort as a control for an older cohort, in order to examine what happens over a definable period of schooling. (I will use the 10-year-old cohort as a control for the 14-year-old cohort, in order to examine differences among countries in schooling at the upper elementary level. I will then use my findings to predict the differing outcomes of schooling that we can expect to find at the secondary level, and I will check my predictions by testing the achievement of students in the various countries during their final year of secondary school.)

• It involves looking at *relative* changes in two areas of achievement in different countries, in order to gain some idea of the relative emphasis that is placed on these two subjects. (The two areas on which I will focus are science achievement and reading comprehension.)

It is important to note that the analysis I am about to describe is not extensive, sophisticated, or highly technical. I have merely used the published data from two volumes of the six-subject IEA study: the volume on science achievement[3] and the volume on reading comprehension.[4]

None of the elements of data analysis that I am suggesting with this example are new for the IEA studies. In the first IEA study of mathematics achievement, Peaker carried out a between-country analysis,[5] and others have been conducted since that time. Younger cohorts have also been used before as controls for older cohorts, in order to obtain measures of growth.[6] Similarly, relative achievement in different subject areas has been examined previously.[7]

However, I maintain that the overall strategy — the combination of the various elements of the data analysis — *is* new. Moreover, this strategy facilitates learning new things about the functioning of educational systems and about the effects of that functioning on students who pass through those systems.

I WILL EXAMINE the use of this strategy in a specific case. In Figure 1 I have plotted the average achievement test scores of 10-year-olds in 10 different countries* in the areas of science and reading comprehension. These average scores indicate the number of items correctly answered out of 40, the total number of items included in each test. (The scores are adjusted to compensate for random guessing.) The data cover all countries — except two less-developed nations, Chile and India — that administered tests in both reading comprehension and science to 10- and 14-year-olds.[8]

*The countries were Sweden (A), Flemish Belgium (B), Finland (C), the U.S. (D), Hungary (E), Italy (F), England (G), the Netherlands (H), Scotland (I), and French Belgium (J). Because of the language difference, French and Flemish Belgium were treated as separate countries.

The scores in Figure 1 show some scatter. Students in Country A scored the highest in both reading and science, but no single country's students scored the lowest in achievement in both areas. Students in Country E, who scored the lowest in reading comprehension, were a little above average in science. Meanwhile, students in Countries I and J scored especially low in science but showed themselves about average in reading comprehension. (It is important to note that, since the 10 countries have eight different languages, the tests of reading comprehension were necessarily different. Although the IEA

"No, Douglas, not in a nanosecond. Come to dinner now!"

went to great lengths to make the reading tests fully comparable, the scores derived from these tests are not as comparable as those derived from the science tests, which covered the same items in all countries.)

Figure 2 uses the scores of the 10-year-olds as starting points for 10 arrows, each of which ends at a point representing the achievement of 14-year-olds in the same country.[9] In all the countries but two, essentially the entire group of 14-year-olds was still in school at the time of testing. In those two countries, 83% and 55% of all 14-year-olds were still in school when the testing took place.

The test scores plotted in Figure 2 do not represent performances on the *same* items by 10- and 14-year-olds. The science test for 14-year-olds consisted of 80 items – twice as many items as were included in the reading test for 14-year-olds and in the reading and science tests for 10-year-olds. Thus the data do not allow any conclusions regarding absolute levels of cognitive growth. Relative growth in the two subjects can be examined, however.

What is striking about Figure 2 are the two different directions taken by the arrows. The arrows for Countries A, E, G, and I are nearly parallel with one another – heading off toward the right, which indicates a higher ratio of cognitive growth in science to growth in reading comprehension. The arrows for Countries B, C, D, F, H, and J are also nearly parallel with one another, but their steeper slope shows a lower ratio of growth in science to growth in reading comprehension than was true for the first set of countries. There are also differences in the overall amount of growth. Country A has the shortest arrow, and Country E has by far the longest. By the time students are 14 years of age, variations in cognitive growth in science have produced a much wider variation in achievement among countries than is true for 10-year-olds. However, the variation among countries in reading achievement is much *less* at age 14 than at age 10.

The striking differences between the two sets of countries, shown by the directions of the arrows in Figure 2, suggest that education between the ages of 10 and 14 differs in some respects for the two sets of countries. The schools in Countries A, E, G, and I appear to be concentrating on science, while those in Countries B, C, D, F, H, and J appear to be focusing on reading comprehension – or, perhaps, more generally on verbal skills or on humanistic education.

What other indicators would provide more evidence? One such indicator is the variation *within* a country in achievement in the subject that appears to be taught more intensively. If the schools are really teaching the subject more intensively, we would expect to find an increased spread among students' test scores in that subject, since some students would progress very rapidly and others would not.

The spread among test scores of 14-year-olds in a given subject can be measured by comparing the standard deviation of those test scores to the standard deviation of the test scores of 10-year-olds in the same subject. (The 10- and 14-year-olds who were tested by the IEA are representative samples of their age groups except in Country E, since nearly all students are still in school at those ages. However, this is not true of the other population tested by the IEA, i.e., students in the last year of secondary school. Countries vary widely in the fraction of the cohort that remains in secondary school until graduation – and thus takes the IEA test. Therefore, we could not compare the standard deviation of scores of, say, 14-year-olds with the standard deviation of scores of secondary school seniors, since the latter constitute differently selected fractions of the cohort in different countries.)

The essential difference in schooling between the two sets of countries appears to be in the intensity with which science is taught.

Table 1 shows the standard deviations of scores of 14-year-olds as multiples of the standard deviations of scores of 10-year-olds. On the left side of that table are the data for Countries B, C, D, F, H, and J (which I will call Group 1) – the countries with a lower ratio of cognitive growth in science to growth in reading. On the right side of Table 1 are the data for those countries that I will call Group 2 (A, E, G, and I) – the countries with a higher ratio of cognitive growth in science to growth in reading.*

The standard deviations of scores of 14-year-olds in reading are only a little greater than those of 10-year-olds, for both sets of countries. In other words, my conjecture that the schools in Group 1 countries are bringing about differential performance in reading comprehension or in general verbal skills through focusing on this area does not seem to be true.

*The countries in Group 1 are Belgium (both French and Flemish), Finland, the U.S., Italy, and the Netherlands. The countries in Group 2 are Sweden, Hungary, England, and Scotland.

Table 1. Standard Deviations of 14-Year-Olds' Scores In Each Country, Relative to Those of 10-Year-Olds

	Group 1 Countries in which growth in science is low, relative to that in reading comprehension		Group 2 Countries in which growth in science is high, relative to that in reading comprehension	
Country	SD 14 / SD 10 Reading	SD 14 / SD 10 Science	SD 14 / SD 10 Reading	SD 14 / SD 10 Science
A			1.03	1.60
B	1.05	1.26		
C	1.01	1.29		
D	1.00	1.25		
E			1.01	1.59
F	1.06	1.19		
G			1.03	1.66
H	1.07	1.32		
I			1.04	1.69
J	0.94	1.24		

Table 2. Average Science Achievement Scores of Top Science Students in Group 1 Countries and in Group 2 Countries

	Country	Score
Group 2	G	51.6
	I	50.7
	A	49.5
	E	48.0
Group 1	H	47.1
	C	46.0
	D	45.8
	B	39.8
	F	38.2
	J	36.2

However, the story with regard to science scores is very different. Those countries with *high* ratios of growth in science to growth in reading comprehension show much greater increases in the spread of students' science scores between the ages of 10 and 14 than do countries in which the ratio of growth in science to growth in reading is lower.

This suggests that the essential difference in schooling between the two sets of countries – for 10- through 14-year-olds, at least – is not in the intensity with which reading and verbal skills are taught, but in the intensity with which science is taught. The countries in Group 2 apparently begin to teach science intensively to at least some of their students by the time the students are 14 years old. The outcome seems to be twofold: the science scores of 14-year-olds in those countries increase overall, and the spread among test scores of 14-year-olds within each of those countries increases.

This conjecture is borne out by the overall increase in science scores and by the increase in spread among students' science scores in two countries, West Germany and Japan, that are not included in the analysis above. (These two countries administered the science test, but not the reading test, to their 10- and 14-year-olds.) In these two countries, the average growth in science was 8.8 and 9.2 respectively – comparable to the average growth of 7.2 that characterized Group 2 countries (whereas Group 1 countries produced an average growth of only 2.7). The standard deviations of the scores of 14-year-olds were also high in these two countries compared to the standard deviations of the scores of 10-year-olds. The standard deviations of the scores of 14-year-olds, expressed as multiples of the standard deviations of scores of 10-year-olds (as in Table 1 for the original 10 countries), are 1.55 and 1.92 for these two countries not included in the original analysis. The standard deviation in Japan, with exceptionally high growth in science, nearly doubled.

A further question comes immediately to mind: What is the later impact on students' achievement of this early focus on science in the schools? Group 2 countries apparently begin to concentrate the attention of at least some of their students on science by age 14, while Group 1 countries do not. Do the Group 2 countries produce an elite corps of students who, by the time they graduate from secondary school, are more proficient in science than their counterparts in Group 1 countries?

The answer appears in Table 2, which lists the average science achievement scores of the top 1% of science students in both Group 1 and Group 2 countries. As Table 2 shows, the scores of the top science achievers in *all* the Group 2 countries are higher than the scores of the top science achievers in *any* Group 1 country.* Clearly, the early focus on – or, perhaps, specialization in – science in the Group 2 countries produces a higher-achieving scientific elite by the end of secondary school. This is a later consequence of a difference that can already be observed when students are only 14 years of age.

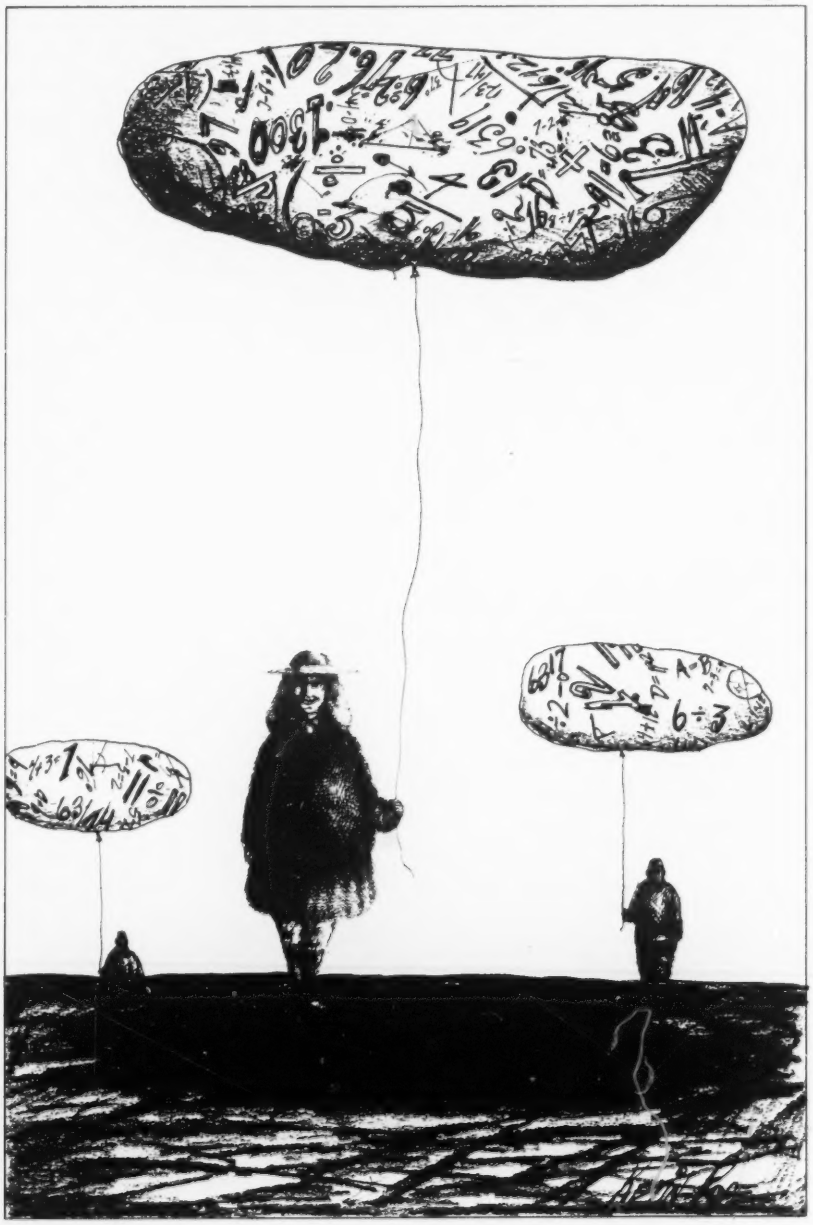
The rather simple analysis that I have carried out here suggests that the IEA data have more potential for providing information on the effects of schooling than has yet been realized. The very design of the IEA project – the fact that it covers educational systems that exhibit a greater range of variation than schools within a single country ever do – makes IEA data a potentially rich source of information on the effects of schooling. Major variations in school systems can significantly affect students' achievement – but to find these major variations, it is necessary to break away from the country-by-country analyses that have typified IEA reports to date.

"We're getting a new computer in our class today. I hope I'm the one it replaces."

*West Germany, one of the two countries that I could not include in my analysis (because it did not administer the reading test), *did* administer the upper-secondary science test. As mentioned earlier, growth in science achievement and dispersion in West Germany are like the countries of Group 2. At the end of secondary school, the average score of the top 1% of science achievers in this country was 45 – below one Group 1 country and above five Group 1 countries.

1. Stephen Heyneman and William Loxley, "The Effect of Primary School Quality on Academic Achievement Across Twenty-Nine High- and Low-Income Countries," *American Journal of Sociology*, vol. 88, 1983, pp. 1162-94.
2. Andrew Greeley, *Catholic Schools and Minority Students* (New Brunswick, N.J.: Transaction Books, 1982).
3. L. C. Comber and John P. Keeves, *Science Education in Nineteen Countries* (Stockholm: Almqvist and Wiksell, 1973).
4. Robert L. Thorndike, *Reading Comprehension Education in Fifteen Countries* (Stockholm: Almqvist and Wiksell, 1973).
5. Torsten Husén, ed., *International Study of Achievement in Mathematics, Vols. 1 and 2* (Stockholm: Almqvist and Wiksell, 1967).
6. See, for example, Thorndike, p. 145; and Comber and Keeves, p. 120.
7. See, for example, Thorndike, Ch. 10, which compares data on reading comprehension, science achievement, and literature achievement from the six-subject IEA study.
8. The scores are taken from Comber and Keeves, p. 159, Table 7.2; and Thorndike, p. 124, Table 8.1.
9. The data are taken from Comber and Keeves, p. 159, Table 7.2; and Thorndike, p. 125, Table 8.2. □

Mathematics Achievement In U.S. Schools: Preliminary Findings From the Second IEA Mathematics Study



by Kenneth J. Travers and Curtis C. McKnight

"More of the same" may not be the kind of reform of secondary mathematics that U.S. students need, say the authors. What to offer and how to teach it are the major issues that confront reformers.

THE NATIONAL spotlight currently focused on the quality of education in the U.S. has heightened interest in the cross-national studies of the International Association for the Evaluation of Educational Achievement (IEA). Now in their second generation, these studies began some 20 years ago with the first IEA study of mathematics, carried out in 12 countries, including the U.S.[1] That study was a forerunner of an even more ambitious round

KENNETH J. TRAVERS is chairman of the International Mathematics Committee and a professor of mathematics education at the University of Illinois, Urbana. CURTIS C. McKNIGHT is an associate professor of mathematics and education at the University of Oklahoma, Norman.

This article is based on the Second International Mathematics Study: U.S. Summary Report. *The major authors of that report also include F. Joe Crosswhite, Ohio State University; Jane O. Swafford, Northern Michigan University; John A. Dossey, Illinois State University; and Thomas J. Cooney, University of Georgia.*

Funding for the Second International Mathematics Study came from several sources, including the National Institute of Education, the National Science Foundation, and the National Center for Education Statistics. The views expressed in this article are those of the authors and do not necessarily reflect the positions of the funding agencies.

Illustration by Kevin Pope

of surveys conducted by the IEA in the late 1960s and early 1970s. These surveys gathered empirical data on achievement from 21 countries in the fields of literature, science, reading, civic education, and English and French as foreign languages.

The first round of the IEA studies received considerable public attention in the U.S. and in many other countries. For example, when the first mathematics study was released, daily newspapers in major cities carried front-page stories and editorials contrasting the accomplishments in mathematics of students in certain countries (notably, Japan) with the relatively poor showing of U.S. students.

However, the IEA studies are *not* designed to be educational Olympiads. Indeed, a recent article by Torsten Husén of Sweden, one of the founders of the IEA, points to the limitations of cross-national comparisons of educational achievement.[2] Husén argues that the highly industrialized countries of western Europe, though in many ways comparable to the U.S., have basically different school systems. For example, the centralized educational systems of Europe, with their classical curricula designed to prepare students for universities, have vastly different objectives from those of U.S. schools, which are locally controlled, have diverse curricula, and are relatively less selective.

Despite their limitations, however, international studies of educational achievement *do* serve important purposes. For example, the international studies compile "benchmark" data on a variety of aspects of education, including achievement data (perhaps the most revealing at the item level), attitude data, and information about such issues as the use of calculators. This information is of use to policy makers and curriculum specialists at national, state, and local levels.

The Second International Mathematics Study, which surveyed 22 countries,* focused on two populations: 1) all students in the grade in which the majority of students have attained the age of 13 by the middle of the school year and 2) all students who are in the normally accepted terminal grade of secondary school *and* who are studying mathematics as a substantial part (approximately five hours per week) of their academic programs.

In the United States, the second population was taken to mean those students who were in college-preparatory mathematics courses requiring as prerequisites at least two years of high school algebra and one year of high school geometry. Although such students were typically in the 12th grade, the definition did allow inclusion in the target group of those students in the 11th grade (and lower) who had moved through the standard sequence of courses more quickly than the average student.

The Study added two aspects of data collection not previously available on a national scale. For a subset of seven countries (Flemish Belgium, France, Japan, New Zealand, Thailand, the United States, and Canada – both British Columbia and Ontario), a "classroom process" component of the Study was carried out among the 13-year-olds. This consisted of a pretest and detailed questionnaires on instructional practices. Thus, for this subgroup of countries, data are available on how much mathematics was learned during the school year and on how the mathematics that was taught during that year was presented by classroom teachers.

The eighth-grade mathematics curriculum in the U.S. resembles a program of studies for the end of elementary school.

*The countries that took part in the Second International Mathematics Study were: Australia, Belgium (French and Flemish), Canada (British Columbia and Ontario), Chile, England and Wales, Finland, France, Hong Kong, Hungary, Ireland, Israel, the Ivory Coast, Japan, Luxemborg, the Netherlands, New Zealand, Nigeria, Scotland, Swaziland, Sweden, Thailand, and the United States.

"I've dedicated my life to mathematics. I've been teaching for (9+3)×2÷4x5 years."

THE REPORTING phase of the Study has now begun, and individual countries are releasing reports on their participation in the Study. The international reports are scheduled to be published in early 1986 by Pergamon Press of London. Even though the international reports are not yet available, the *U.S. Summary Report* offers useful information on the teaching and learning of mathematics in U.S. schools, as well as some preliminary indicators of how the U.S. stands in relation to other countries.

The timing of the Study was fortunate. Much has taken place since the first study was conducted in 1964. In the early 1960s curriculum reform in mathematics was just getting under way. Since then, considerable curriculum development has taken place, and varieties of educational technology have appeared. Now that we are making decisions about directions for the 1990s and beyond, it's a good time to get another measure of our progress in relation to other countries.

The planners of the Second IEA Mathematics Study had the broad concerns of mathematics educators clearly in mind when they formulated the project. Plans for the Study were developed at an international meeting held at the University of Illinois in 1976. These concerns clustered in three main areas: the curriculum, the

classroom, and the student. Or, put more formally, the *intended* curriculum, the *implemented* curriculum, and the *attained* curriculum.

The *intended* curriculum is reflected in curriculum guides, course outlines, syllabi, and textbooks adopted by school systems. In most countries, national curricula emanate from a ministry of education or from a similar body. In the U.S., of course, such statements of intended goals and curricular specifications come from state departments of education and from local districts. Thus it is considerably more difficult to describe the intended curriculum for the U.S. than for almost any other country that took part in the IEA Study.

The *implemented* curriculum focuses on the classroom, where the teacher interprets and puts into practice the intended curriculum. Teachers exercise their own judgment in translating curriculum guides and adopted textbooks into programs for their classes. Hence, their selection of topics or patterns of emphasis may not be consistent with the intended ones.

To identify the implemented curriculum, a number of questionnaires were developed for classroom teachers to complete. For example, teachers were asked whether or not they had provided instruction for each of the items on the achievement tests. They were asked questions about such topics as the use of calculators in their classes. They were also asked to provide detailed information on the number of class periods that they devoted to specific topics and subtopics and on how they presented and interpreted them to their classes.

For example, one of the questionnaires dealt with the ways a teacher could interpret fractions. These interpretations ranged from the concrete (fractions as measurements or fractions as regions) to the abstract (fractions as decimals or fractions as quotients). The questionnaire asked teachers whether they interpreted fractions for their students in one of 10 ways and whether they used a particular interpretation occasionally or frequently. Such highly specific information on classroom presentations, coupled with similarly detailed information on instructional strategies, permits us to draw a rather comprehensive picture of *what* and *how* mathematics has been taught to specific populations.

On the use of calculators, we found that, in the year in which the data were collected (1981-82), calculators had found their way into relatively few eighth-grade mathematics classrooms. In about one-third of the classrooms, calculators were not even allowed. Another one-third reported that they were never used. In most of the remaining one-third, calculators were used only one period or less per week.

In the 12th grade, calculators were more commonplace. About one-third of the teachers reported using calculators two or more periods per week, almost another one-third reported using calculators one period or less per week. Only 10% of teachers reported that their students were not allowed to use calculators at all. The most commonly reported instructional uses of calculators in the 12th grade were checking work, doing homework, and solving problems in class. About 50% of 12th-graders were allowed to use calculators on tests.

The *attained* curriculum – what students have learned as measured by tests and questionnaires – was the third (and most familiar) component of the Study. Extensive achievement tests were designed to assess student knowledge and skills in areas of mathematics that were designated as important and appropriate for each population. The "fit" between these tests and the actual curricula in individual countries may vary considerably, because the tests contain items that are less appropriate in some countries than in others and because they could not possibly contain an adequate range of items to *fully* represent all curricula in all countries. This reduces the accuracy of cross-national comparisons, and this limitation provides another argument against interpreting the Study as an international contest.

The student outcome measures also included a number of opinionnaires and attitude scales. These were devised to elicit students' views on the nature, importance, ease, and appeal of mathematics in general and of selected mathematical processes.

Each country taking part in the Study drew a national probability sample in accordance with an international sampling design. The procedures were monitored by an international sampling referee. In the case of the Second Mathematics Study, Malcolm Rosier of the Australian Council for Educational Research served as the sampling referee. The total U.S. sample consisted of students and teachers from approximately 500 classrooms in public and private schools across the country.

The students then took internationally developed mathematics achievement tests at the beginning of the 1981-82 school year; attitude and general background questionnaires were administered at the same time. Some 7,000 eighth-grade students and some 4,500 12th-grade students from approximately 250 public and private schools took part, along with their teachers.

EIGHTH-GRADE FINDINGS

Intended curriculum. From the perspective of other nations, the eighth-grade mathematics curriculum in the U.S. looks more like a program of studies for the end of elementary school than for the beginning of high school. Our curriculum appears to be dominated by topics in arithmetic and measurement rather than by the study of algebra and geometry.

Implemented curriculum. One clear finding of the Study is that there is great diversity in the mathematics taught in U.S. schools at the eighth-grade level. Four distinct types of classes were identified: 24 remedial classes, 26 enriched classes, 31 algebra classes (studying only algebra for a full year and using an algebra text), and 155 "typical" classes.

What mathematics content was taught in these classrooms? For each of the 180 items on the achievement tests, teachers were asked to indicate whether the content needed to correctly respond to an item was taught during the eighth grade, was taught in a previous year, or was not taught at all. Table 1 presents the content coverage during eighth grade for five topics (groups of test items) and four class types.

Notice that typical and enriched classes covered a large portion of the content in all areas. Enriched classes covered more than typical classes in algebra, geometry, and measurement. Remedial classes emphasized arithmetic and covered less algebra and geometry. The algebra classes covered a considerable amount of algebra (actually more than was on the test), but they covered little else. No class covered much of the geometry that was on the international test, which included transformational geometry (not usually taught at all in the U.S.) and spatial visualization.

The detailed classroom-process questionnaires showed that the teaching of

Table 1. Percent of Items Taught During Eighth Grade

		Class Type			
Content Area	**N**	**Remedial**	**Typical**	**Enriched**	**Algebra**
Arithmetic	62	76	80	84	42
Algebra	32	37	64	80	86
Geometry	42	25	41	54	24
Measurement	26	53	64	75	20
Statistics	18	48	58	59	15
Overall	180	51	64	72	40

common and decimal fractions tended to be symbolic and formal rather than intuitive and concrete. The teaching of ratio and percent also tended to be formal. When teachers presented interpretations of these concepts, they were typically formal as well. For example, ratio was often interpreted as a fraction, but fractions were themselves presented through formal and symbolic means. However, integers were often introduced through the number line and through physical situations. Still, when they presented operations using integers, teachers relied on giving rules.

Topics in geometry were typically introduced by a statement of definitions and properties, rather than by intuition or informal explanation. For example, the property of the sum of the angles in a triangle was usually taught by presenting and demonstrating it, or by having students find the sum of the measures of the angles in various triangles. Paper folding and other concrete activities were not emphasized.

In geometry, teachers' expressed opinions were greatly at odds with their reported practices. Although teachers affirmed that intuitive approaches are most meaningful for students, that concrete models and aids have instructional value, and that spatial visualization is important, they reported that they used primarily formal, "tell and show" approaches to instruction. Data on how students spend their time appear to support this finding. The majority of student time in mathematics classes was reported to be spent in listening to teacher presentations, doing seatwork, or taking tests.

Attained curriculum. Table 2 presents a summary of mean eighth-grade achievement scores for each of the five content areas — arithmetic, algebra, geometry, statistics, and measurement. Separate scores for both the pretest and posttest are listed.

The column headed OTL (opportunity to learn) lists the average rating by teachers of the percent of content that had been taught either before or during the year. This figure gives some indication of how well the tests matched the U.S. curriculum. These ratings represent a sort of "ceiling" on how well U.S. students can be expected to do. Notice that geometry has the lowest OTL score.

Figure 1 shows a sample test item from the algebra portion of the international test for eighth-graders. The OTL rating of 75 indicates that three-fourths of the teachers in the sample reported that their classes had been taught the mathematics needed to answer this item correctly either during the eighth grade or earlier. Sixty percent of students nationwide answered this question correctly at the beginning of the year, and 71% answered it correctly at the end of the year. Internationally, the median posttest score was 67% correct.

Table 3 compares mean end-of-year achievement for the U.S. eighth-grade sample with the 25th percentile, median, and 75th percentile scores of all participating countries combined. In algebra, arithmetic, and statistics, the U.S. performed at about the international median. In geometry and measurement, the U.S. performed at about the 25th percentile. In no case was the U.S. performance above the international median.

Table 2. National Estimates of Mean Scores In Eighth-Grade Classes

Content Area	N	OTL*	Mean Scores		
			Pretest	Posttest	Gain
Arithmetic	62	87	42	51	+ 9
Algebra	32	69	32	43	+11
Geometry	42	44	31	38	+ 7
Measurement	26	70	35	42	+ 7
Statistics	18	73	53	57	+ 4

*OTL = opportunity to learn.

Figure 1. Sample Eighth-Grade Algebra Item

A bowling ball travels four meters per second. The distance in meters traveled in t seconds is given by $d = 4t$. In the table below x is equal to:

t	d
0	0
1	4
2	8
3	x
4	16

A. 6
B. 10
C. 12
D. 14
E. None of these

	% Correct	
OTL*	Pretest	Posttest
75	60	71
International Median		67

*OTL = opportunity to learn.

Table 3. International Comparisons of Eighth-Grade Achievement

Content Area	N	U.S. Mean (% correct)	25th %ile	International Median	75th %ile
Arithmetic	62	51	45	51	57
Algebra	32	43	39	43	50
Geometry	42	38	38	43	45
Measurement	26	42	47	51	58
Statistics	18	57	52	57	60

12TH-GRADE FINDINGS

Intended curriculum. In the 12th grade in 1981-82, we estimated that about 12% of our students were enrolled in college-preparatory mathematics courses. The curriculum in these classes is predominantly composed of topics regarded as preliminary to the study of calculus, such as trigonometry, analytic geometry, and elementary functions.

Although calculus itself may be taught in the 12th grade in U.S. schools, it does not receive as much emphasis in the U.S. as it does in many other nations —

notably, Japan and several European countries. Moreover, the proportion of U.S. students enrolled in full-fledged calculus courses is lower than in most other countries taking part in the Study. This is true even when the relatively lower retention rates of school systems in other countries are taken into account. Although other countries may keep smaller proportions of students in school through the end of the secondary program, those students who are still studying college-preparatory mathematics are almost universally studying calculus.

The data are not yet final and will be made available in the international reports. However, we are estimating that about 2% to 3% of U.S. students graduate from high school with a solid introduction to calculus, while in Europe at least 5% to 6% of students who have completed a college preparatory mathematics program have received a solid introduction to calculus. In Japan, that figure rises to about 12%.

Implemented curriculum. Two types of 12th-grade classes were identified: calculus classes (defined as those that followed the College Board's Advanced Placement calculus program) and precalculus classes that studied various combinations of trigonometry, analytic geometry, elementary functions, and introductory topics in calculus. In the national sample of 237 classes, 46 were defined as calculus classes and 191 as precalculus classes.

Table 4 shows the content that teachers reported had been covered either during the 12th grade or earlier. The seven content areas are those used on the subtests of the international test. Note that teachers estimate that the precalculus classes have covered about the same amount of content as the calculus classes in the first four areas (either during or prior to the 12th grade). The calculus classes show a more specialized pattern in the last three content areas, particularly in the study of elementary functions and calculus.

The teachers noted that most of the emphasis in teaching analytic geometry was placed on developing students' two-dimensional skills, especially in conic sections and curve sketching. Very little, if any, work was done on solid analytic geometry. In teaching trigonometry, little time was devoted to inverse functions, their graphs, or the relationships involved. The same pattern of relative neglect holds for trigonometric relationships involving representations of complex numbers or the use of polar coordinates.

As was the case in analytic geometry, material dealing with three dimensions was rarely taught in calculus. The emphasis was on the development of limits, differentiation, and integration through logarithmic and exponential functions. Teachers reported little work involving multi-variable calculus, nor was there much emphasis on special techniques of integration, numerical integration, number series, or related topics.

The textbook stood out as the most commonly and consistently used resource in the 12th grade, used by 95% of the teachers. Self-written tests were used by 94% of the teachers, and other self-written materials were used by 60%. The rating for the next most frequently used instructional resource (12% for workbooks and supplementary texts) shows just how narrow is the range of instructional materials used by most teachers.

Attained curriculum. Table 5 presents a summary of mean 12th-grade student achievement scores for each of the seven content areas. Separate scores for both the pretest and posttest are listed for precalculus and calculus classes. The OTL scores in Table 5 represent the mean teacher ratings of the percent of content to which students had been exposed by the end of 12th grade.

Note that the calculus classes made very large gains in elementary functions and calculus — the content area in which most of their efforts were concentrated. However, they made few gains (and even incurred some losses) in other areas. Also note that in all cases the posttest scores of the precalculus classes were below the pretest scores of the calculus classes. However, the precalculus classes did achieve modest gains in all content areas.

Figure 2 presents a sample item from the 12th-grade test and shows how U.S. students performed on this item. The calculus classes had a greater opportunity to learn the content of this item than did the precalculus classes. Consequently, the higher achievement should not be unexpected. The international median happens to correspond exactly to the performance of the calculus classes in the U.S.

Table 4. Percent of Items Taught in 12th-Grade Classes

		Precalculus		Calculus	
Content Area	N	Taught Before	Taught This Year	Taught Before	Taught This Year
Sets and Relations	7	31	50	50	40
Number Systems	17	39	42	75	14
Algebra	26	34	52	53	41
Geometry	26	21	40	41	26
Elementary Functions and Calculus	46	8	37	9	83
Probability and Statistics	7	29	14	50	6
Finite Mathematics	4	29	21	62	8

Table 5. National Estimates of Mean Scores in 12th-Grade Classes

		Precalculus				Calculus			
Content Area	N	OTL*	Pretest	Posttest	Gain	OTL*	Pretest	Posttest	Gain
Sets and Relations	7	81	48	54	+6	90	66	64	− 2
Number Systems	17	81	33	38	+5	69	43	48	+ 5
Algebra	26	87	35	40	+5	94	53	57	+ 4
Geometry	26	60	24	30	+6	67	35	38	+ 3
Elementary Functions and Calculus	46	45	18	25	+7	92	26	49	+23
Probability and Statistics	7	43	36	39	+3	57	48	48	0
Finite Mathematics	4	50	24	29	+5	71	36	38	+ 2

*OTL = opportunity to learn.

Table 6 compares the mean posttest achievement of U.S. 12th-graders – as a whole and for each class type – with the standardized scores of all nations that took part in the Study. Overall, the U.S. sample performed at a level markedly below the average level of performance found in the final year of secondary school mathematics in the other nations. Considered separately, the calculus classes performed at or near the international median in almost all content areas; the precalculus classes never reached a comparable level.

CHANGES OVER TIME

Table 7 shows the achievement scores for eighth-graders in both the First and Second International Mathematics Studies. The data indicate that overall there has been a modest decline in eighth-grade achievement between 1964 and 1982. While students held their own in algebra, marked declines took place in arithmetic and geometry. Further analysis (not shown here) indicates that much of the decline stems from poorer performance on items requiring higher cognitive skills, rather than on items requiring computational skills.

Table 8 compares the scores of 12th-graders in 1964 with those of 12th-graders in 1982. Overall, there has been a modest gain in achievement for the college-preparatory mathematics students over the 20 years between the two assessments. The increase of six percentage points in elementary functions and calculus is worthy of note, but that gain comes almost entirely from the calculus classes. An analysis of the score changes by cognitive levels (not shown here) indicates that much of the gain stems from items at higher cognitive levels (e.g., comprehension and application), rather than from items requiring computational skills.

STUDENT ATTITUDES

Overall, the attitudes of U.S. students toward mathematics tend to be positive. Students feel that the study of mathematics helps them to think logically and that the subject is a good one for creative people. They also seem to have positive self-concepts with regard to mathematics, and the majority report that both they and their parents want them to do well in the subject.

Students feel that the study of mathematics helps them to think logically and that the subject is a good one for creative people.

Moreover, we did not find evidence of sex-role stereotyping. U.S. students do not generally believe that mathematics as a field of study is a male domain. Finally, students have positive attitudes about the use of calculators in learning mathematics and about the role of computers in society.

It is important to remember, though, that the 12th-grade sample of students was a specialized group – those engaged in the study of mathematics throughout high school. These are students who have been at least somewhat successful in the subject. Hence, positive attitudes should be expected from this group. The fact that the eighth-graders – a cross section of 13-year-olds nationwide – also had positive attitudes about mathematics is more encouraging.

THE FINDINGS of the Second IEA Mathematics Study that we have reported here are preliminary. Although they lack a certain amount of detail, they still provide a unique vantage point from which to view

Figure 2. Sample 12th-Grade Algebra Item

Two mathematical models are proposed to predict the return y, in dollars, from the sale of x thousand units of an article (where $0 < x < 5$). Each of these models, *A* and *B*, is based on different marketing methods.

Model *A* $y = 6x - x^2$
Model *B* $y = 2x$

For what values of x does model *B* predict a greater return than model *A*?

A $0 < x < 4$
B $0 < x < 5$
C $3 < x < 5$
D $3 < x < 4$
E $4 < x < 5$

Type of Class	OTL*	% Correct (Posttest)
Precalculus	54	45
Calculus	80	56
Combined		47
International Median		56

*OTL = opportunity to learn.

Table 6. International Comparisons of 12th-Grade Achievement

Content Area	N	U.S. Means: Precalculus (% correct)	U.S. Means: Calculus (% correct)	U.S. Means: Combined (% correct)	International Scores: 25th %ile	International Scores: Median	International Scores: 75th %ile
Sets and Relations	7	54	64	56	51	61	72
Number Systems	17	38	48	40	40	47	59
Algebra	26	40	57	43	47	57	66
Geometry	26	30	38	31	33	42	49
Elementary Functions and Calculus	46	25	49	29	28	46	55
Probability and Statistics	7	39	48	40	38	46	64
Finite Mathematics	4	29	38	31	--	44*	--

*Estimated median.

the teaching and learning of mathematics in U.S. schools. And they raise questions about how the current proposals to reform secondary education should deal with mathematics.

An analysis of the content coverage in eighth grade reveals a great diversity. For example, a full range of topics in geometry was covered in the U.S. sample. A few classes reported that they included no geometry, a few included nearly all the geometry on the international test, and most classes fell somewhere in between.

This diversity can be attributed to our decentralized system of education, in which specific curriculum decisions are left to local school boards or to classroom teachers. Our freedom from a centralized system clearly offers the opportunity for local authorities to provide students with the kind of program that they judge to be most appropriate. However, this freedom is purchased at a price. That price is the difficulty in obtaining a national commitment to agendas for upgrading instruction.

At the eighth-grade level, where the curriculum appears to be an extension of the elementary school – rather than the beginning of the high school, as is the case in a number of other countries – the Study raises even more fundamental questions. For instance, the differentiation in curriculum that takes place in eighth grade is troublesome. While a few students are already studying high school algebra, a good many others are funneled into remedial classes and may even be effectively cut off from the study of high school mathematics. At the same time, most students are in regular eighth-grade classes, studying arithmetic and not receiving sufficient preparation for the study of high school mathematics.

The findings of the Study also raise some fundamental questions at the 12th-grade level. First, the calculus classes that follow the Advanced Placement (AP) regimen are giving students a solid introduction to calculus. On the average, U.S. calculus students compare well with those in many other countries. The critical difference is that only a small proportion of U.S. students are in such programs. Data from other sources indicate that only about 35,000 students took the AP calculus tests in 1982. Although this number has been increasing rapidly over the past 20 years, it still represents only a small percentage of the students in senior high schools.

A second fundamental question that the Study raises involves the kind of mathematics that is being taught to high school seniors. Our data indicate that most 12th-grade mathematics is designed to prepare students for calculus. A debate is now under way about how relevant such a program is for most students.

The recommendations of the national commissions and task forces, if they are heeded, would increase the amount of mathematics required for high school graduation. But the data from the Second IEA Mathematics Study indicate that many students have not reached the level of achievement needed to deal with college-level calculus. Indeed, "more of the same" may not be the kind of reform of secondary mathematics that U.S. students need. The National Science Board Commission has stated that, as the effects of computers and other high technologies come to be felt in society, discrete mathematics (including statistics and probability) should be regarded as "fundamental for all students."[3] The kind of mathematics to be offered to those students who will be taking more mathematics than ever before and how to teach it to them are among the major issues that confront those who would reform mathematics education in the U.S.

Data from the study indicate that many students have not reached the level of achievement needed to deal with college-level calculus.

Table 7. Eighth-Grade Achievement on First and Second International Mathematics Studies

Content Area	N	First Study (% correct 1964)	Second Study (% correct 1982)	Change
Arithmetic	14	55	49	−6
Algebra	10	40	41	+1
Geometry	5	40	34	−6
Statistics	5	57	54	−3
Measurement	2	35	37	+2
Overall	36	48	45	−3

1. Torsten Husen, ed., *International Study of Achievement in Mathematics*, Vols. 1 and 2 (New York: Wiley, 1967). Twelve nations took part in the First IEA Mathematics Study: Australia, Belgium (French), the Federal Republic of Germany, England, Finland, France, Israel, Japan, the Netherlands, Scotland, Sweden, and the United States.

2. Torsten Husén, "Are Standards in U.S. Schools Really Lagging Behind Those in Other Countries?," *Phi Delta Kappan*, March 1983, pp. 455-61.

3. National Science Board Commission on Precollege Education in Mathematics, Science, and Technology, *Educating Americans for the 21st Century* (Washington, D.C.: National Science Board, 1983), p. 43. □

Table 8. 12th-Grade Achievement on First and Second International Mathematics Studies

Content Area	N	First Study (% correct 1964)	Second Study (% correct 1982)			Change
			Precalculus	Calculus	Combined	
Sets and Relations	1	21	52	70	56	+35
Number Systems	1	45	56	72	59	+14
Algebra	3	44	43	59	45	+ 1
Geometry	4	36	36	45	37	+ 1
Elementary Functions and Calculus	10	25	27	50	31	+ 6
Probability and Statistics	1	45	48	56	49	+ 4
Overall	20	32	35	53	38	+ 6

The Second International Science Study: U.S. Results

by Willard J. Jacobson and Rodney L. Doran

Among the surprising findings of the Second International Science Study is the fact that fifth- and ninth-graders in the U.S. showed a pervasive improvement between 1970 and 1983. The results of the Study are useful to all policy makers in education.

THE U.S. IS one of 25 nations currently engaged in the Second International Science Study (SISS), which is being conducted under the auspices of the International Association for the Evaluation of Educational Achievement (IEA). The other nations are: Australia, China, Canada, England, Finland, Ghana, Hong Kong, Hungary, Israel, Italy, Japan, the Republic of Korea, Mexico, the Netherlands, Nigeria, Norway, Papua New Guinea, the Philippines, Poland, Singapore, Sweden, Tanzania, Thailand, and Zimbabwe. With 25 participants, this is one of the largest international empirical studies ever undertaken.

Only data on the U.S. are available to us at this time. Data collection in some nations is not yet complete, and no international comparative data will be available until 1986, at the earliest. Nevertheless, the preliminary results on science achievement among U.S. students are interesting and potentially useful to those who make education policy.

Four populations of U.S. students were studied, each made up of intact classes selected at random in 125 schools chosen to be representative of those four populations nationwide. The sample was stratified so that about 90% of the schools were public and 10% were nonpublic. Students fell into the following population categories: fifth grade, ninth grade, 12th-grade students taking physics, and 12th-grade students not taking science.

The students in these four groups were given three separate tests during three separate testing periods. The first test was the "core" international test used in all the participating countries. For fifth- and ninth-graders, the second test contained rotated forms of subtests, including national option items. For the 12th-grade physics population, the second test was an international physics test that also included five U.S. national option items. The 12th-grade students who were not studying science took tests that contained items related to science and technology that could be useful to future citizens. The third test given to each of the populations consisted of four parts: a questionnaire requesting specific information about the students, a student opinionnaire (SISS Attitude Scales), a science learning inventory, and a word knowledge test.

There was also a school questionnaire, to be completed by the local test coordinator or by an appropriate school official. This questionnaire requested information about the characteristics of the school, the community it serves, the organization of the school day and week, the patterns of science instruction, the physical facilities, and the way in which decisions related to education are made in the school and community.

A questionnaire given to the teachers contained questions about them and their backgrounds. The teachers were asked how they decided what to teach and how they decided which approaches to teaching and testing to use. For each test item, the teachers were also asked to indicate whether or not their students had had a chance to study the concepts tested.

The SISS data for U.S. students have revealed some surprising findings, which should be carefully considered in the formulation of science education policy. In this article, we will emphasize the results for students in grades 5 and 9.

ONE USEFUL feature of the SISS is that the data allow us to compare the achievement of U.S. students in 1983 with the performance of their counterparts in 1970, when the First International Science Study was conducted. A number of items from the first study — called "bridge" items — were used in the SISS, and the scores on these items are of special interest.

At the fifth-grade level, there were 26 bridge items. Overall, students in 1983 did significantly better on these items (their scores were 3.5% higher) than students tested in 1970. On 18 bridge items, the means in 1983 were more than 2% higher than in 1970. On five items there was no significant difference, and on three items students in 1970 outscored students in 1983.

At the ninth-grade level, there were 33 bridge items. Overall, students in 1983 performed significantly better (their scores were 2.8% higher) than students tested in 1970. The means of 15 items were more than 2% higher in 1983 than in 1970. On 11 items the differences were less than 2%; on seven items, students in 1970 did better than their counterparts in 1983.

At the fifth-grade level, students had higher scores on bridge items that were classified as physical science than on those classified as biology. But the gains from 1970 to 1983 were greater in biology than in physical science. The fifth-graders in 1983 answered about 59% of the biology items correctly, and this represented a

WILLARD J. JACOBSON is national research coordinator for the Second International Science Study (SISS) and a professor of natural sciences at Teachers College, Columbia University, New York, N.Y. RODNEY L. DORAN (State University of New York/Buffalo Chapter) is a professor of science education at the State University of New York at Buffalo and associate national research coordinator of the SISS. *The authors wish to thank the Spencer Foundation for its support of this research.*

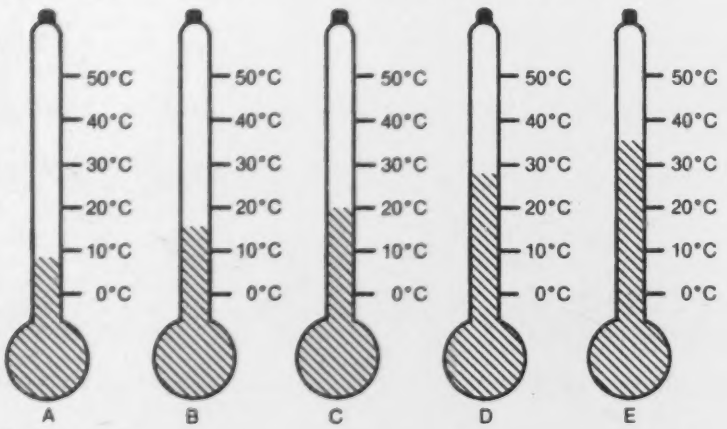
Table 1. Performance on Bridge Items by Content Area

Grade 5	1970 (% correct)	1983 (% correct)	Net Gain (%)
Biology (9 items)	52.8	58.9	+6.1
Physical Science (17 items)	62.2	64.4	+2.2
Total (26 items)	59.0	62.5	+3.5
Grade 9			
Biology (10 items)	62.3	65.6	+3.3
Physical Science (23 items)	54.2	56.8	+2.6
Total (33 items)	56.7	59.5	+2.8

Figure 1. An Item from the Grade 5 Test, Requiring Process Skills

	6 a.m.	9 a.m.	12 noon	3 p.m.	6 p.m.
Monday	15°C	17°C	20°C	21°C	19°C
Tuesday	15°C	15°C	15°C	10°C	9°C
Wednesday	8°C	10°C	14°C	14°C	13°C

Which of the following shows the temperature at 6 a.m. on Wednesday?



	A	B	C	D	E	OMIT
1970	53.3	20.1	8.9	5.0	10.6	2.1
1983	64.6	14.6	8.9	3.9	7.5	.6

gain of 6.1% over 1970. The students answered 64.4% of the physical science items correctly, and this represented a gain of 2.2% over 1970 (see Table 1).

At the ninth-grade level, students had higher scores on bridge items classified as biology than they did on those classified as physical science. In addition, the gains from 1970 to 1983 were slightly greater in biology than in physical science. The ninth-graders in 1983 answered 65.6% of the biology questions correctly, and this represented a gain of 3.3% over 1970. They answered 56.8% of the physical science items correctly, and this represented a gain of 2.6% over 1970.

One of the most surprising findings of the preliminary analyses of the test results for fifth- and ninth-graders in the U.S. is the pervasive improvement in scores for 1983 over scores for 1970. Gains occurred on 18 of 26 items at the fifth-grade level and on 15 of 33 items at the ninth-grade level. Why such improvement between 1970 and 1983?

One way to try to answer this question is to look at the nature of the test items. Many of the items on which scores improved assessed higher levels of thinking and such process skills as classification, calculation, measurement, and the analysis of data. At the fifth-grade level, for example, students did better in 1983 on an item requiring interpretation of tabular data. Figure 1 shows another item on which fifth-graders in 1983 did better than fifth-graders in 1970. This item requires reading tabular data and then applying it to the reading of thermometer scales. As another example of improved process skills, Figure 2 shows an item on which ninth-graders in 1983 were better able (45.6% to 31.3%) to infer the nature of the electrical circuit enclosed within a "black box."

The data suggest that U.S. students in 1983 were better able to respond to items dealing with process skills. In most of the newer science programs, the development of process skills and the development of higher levels of thinking have been emphasized over the recall of factual material. Does this mean that our new emphasis on process skills has been a success? Do these results indicate directions for future improvement?

In 1970, U.S. students found an item dealing with biological adaptation harder than students in all other countries except England and Scotland,[1] but they registered significant gains on this item in 1983. Were these gains due to increased emphasis on the teaching of ecological topics and fewer restrictions on the teaching of topics related to evolution?

These improved test scores in science may also constitute one bit of evidence for a more general improvement in education.

Scores on the Scholastic Aptitude Test have now gone up for three consecutive years.[2] The gains in science could thus reflect a broader trend in U.S. education.

Considering the items on which the scores of U.S. students declined between 1970 and 1983 is also revealing. All three items on which there were declines in the fifth grade and four of the seven items on which there were declines in the ninth grade dealt with topics in earth and space science. Two items that dealt with the length of time it would take a rocket ship to reach the moon and with the nature of moonlight were used as bridge items in both the fifth and ninth grades.

It is well to recall that 1969 and 1970 were years that saw humankind's first ventures to the moon through the Apollo Project. Descriptions and discussions of exploring the moon were everywhere, including on television. Students in 1970 witnessed the blast-off of the rockets and experienced firsthand the wait of several days before the lunar modules landed. Students in 1983 had no such experience. Although the nature of moonlight is mentioned in many textbooks and science programs, there was certainly more discussion and explanation of the phenomenon in 1969 and 1970 than there was in 1983. Thus the decline in these earth and space science items could well be related to the influence of the popular culture, especially television coverage, on the education of U.S. students.

MALES ACHIEVED higher scores than females in the First International Science Study, and the differences between the sexes increased as the level of education increased. In the SISS, differences between the scores of males and females were also evident. Table 2 shows that males scored higher than females on the bridge items at the fifth- and ninth-grade levels and on the core test and the physics test at the 12th-grade level. These gender-related differences widened from grade 5 to grade 9 and from grade 9 to grade 12 – except for the 12th-grade physics population. Apparently, females who enroll in physics (approximately 34% of the physics sample) do better in science than females in the other population groups, and the difference between males and females in the physics population is less than the difference in the total 12th-grade sample.

These findings are generally consistent with those reported in other studies. In the report of the First International Science Study, L. C. Comber and John Keeves commented, "For those who believe that girls are not given fair treatment, the IEA findings provide dramatic evidence of the scope of the problems."[3]

Similar findings were also reported by the National Assessment of Educational Progress in 1981: "In summary, there remains a statistically significant difference between males and females, as well as between white and black students, for the content area."[4] Obviously, these problems still exist, and nagging questions remain. What are the explanations for these differences between boys and girls? What can we do to improve the science achievement of girls? A closer look at the items on which the differences between the sexes were greatest might suggest some answers.

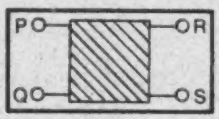
There were 72 items in the fifth-grade test. Fifth-grade boys did better than fifth-grade girls on 53 items; girls did better on only four; and on 14 items there was no significant difference. The fifth-grade girls did better on biology items focusing on a balanced diet and on knowledge about seeds. Girls also outperformed boys on physical science items that asked them to suggest a reason why "a rubbed comb picks up small bits of paper" and to suggest "procedures to use when studying rocks brought from the moon."

The items on which boys did better were primarily in the physical sciences. In the physical sciences, boys did better than girls on 21 items; girls did better on one; and there were no significant differences on six items. Figure 3 shows the item on which there was the greatest difference between boys and girls; 60.9% of the boys answered this item correctly, compared to only 34.5% of the girls. Boys also did better on items assessing the correct placement of batteries in a flashlight, the comparative electrical conductivity of several substances, the reason that a thrown ball comes back to the ground, and how long candles would burn in sealed boxes of different volumes. In addition, more boys knew where fish get the air they need to live. These findings suggest that experiences with physical materials, such as balls, flashlights, and aquariums, might be beneficial for girls.

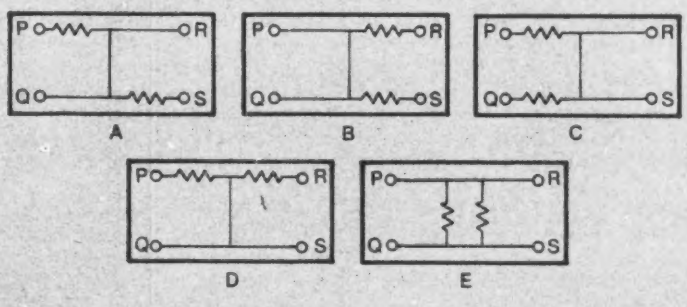
Figure 2. An Item from the Grade 9 Test, Requiring Higher Levels of Thinking

The figure shows a box with four terminals: P, Q, R, and S. The following observations were made.

1. There is a certain amount of resistance between P and Q.
2. Resistance between P and R is twice that between P and Q.
3. There is not any appreciable resistance between Q and S.



Which of the following circuits is most likely to be within the box? Assume that the resistances shown are equal.



	A	B	C	D	E	OMIT
1970	11.3	9.2	11.3	31.3	30.1	6.7
1983	8.5	6.3	14.1	45.6	23.8	1.6

Among ninth-graders, the differences between the sexes were even greater than among fifth-graders. On the 90 items at the ninth-grade level, the boys scored significantly better than the girls on 65; the girls scored better on only three; and there were no significant differences on 22 of the items.

As at the fifth-grade level, the items on which the females had higher scores were in the area of biology. The three items on which ninth-grade females did better assessed "which organ of several is not situated in the abdomen," "which human characteristics are probably due to heredity," and "how to set up an experiment about the germination of seeds." The items on which the boys scored a great deal better than the girls were all in the area of physical science. Three of these items were in earth science; they asked about the "nature of moonlight," "the positions of the sun, the earth, and the moon during a solar eclipse," and "the position of the sun in the morning." Two of these three items called on abilities in spatial perception. The other physical science items on which boys scored a great deal higher than girls measured "the tension between balls hanging on a string," "the need for air in order to communicate by talking," "which objects will *not* be picked up by a magnet," "what happens to light passing through a magnifying glass," and "why evaporated and distilled salt water does not taste salty."

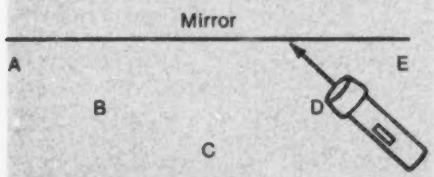
On the 12th-grade core test, boys did significantly better than girls on all items but one, and on that single item there was no significant difference. The item on which there was the greatest difference between the scores of boys and girls was a traditional physics item that involved calculating the tension on a thread supporting objects of different masses. These results seem to suggest the importance for girls of having firsthand experiences in handling such concrete materials as those used in laboratory-style science instruction. Some experience in spatial projection might also be useful.

Table 2. Test Performances by Population and Gender

		% Correct	% Difference
Grade 5	Male	61.3	5.7*
	Female	55.6	
Grade 9	Male	64.7	6.4*
	Female	58.3	
Grade 12 Core Test	Male	62.0	10.6*
	Female	51.4	
Grade 12 Physics Test	Male	37.4	4.8*
	Female	32.6	

*The *t*-tests were significant at the .01 level.

Figure 3. The Item from the Grade 5 Test on Which The Difference Between the Sexes Was Largest



Your friend shines a flashlight at a mirror, as shown above. Where should you stand so that the light will be reflected onto you?

A Position A
B Position B
C Position C
D Position D
E Position E

	A	B	C	D	E	OMIT
Boys	4.5	60.9	22.4	7.7	2.6	2.0
Girls	5.7	34.5	37.8	13.9	4.7	3.4

The SISS is an international study. When the testing has been completed in all 25 countries and the preliminary analyses have been carried out at the International Center in Australia, it will be possible to make international comparisons. Though the comparisons of total scores will be interesting, detailed probing of the item-by-item results may be more useful. We will want to know whether there have been gains between 1970 and 1983 in other countries and whether these gains were on items that deal with the processes of science.

Other information in the SISS will also be of interest. For example, it has been suggested that girls do better in science when there are more female science teachers. In one of the countries participating in the SISS most of the science teachers are women. How will the scores of boys and girls compare in that country?

Another valuable feature of the SISS is its explanatory dimension. An attempt will be made to find explanations for individual findings. A number of intriguing questions suggest themselves: What factors in schools in different countries are associated with achievement in science? Are certain teacher characteristics associated with science achievement? Are there features of science education associated with science achievement in other countries that are not present in U.S. schools?

By asking and attempting to answer such questions, the Second International Science Study represents far more than the international comparisons of total achievement that receive so much coverage in the media. The SISS has the potential for discovering ways to improve the science education of our children and young people.

1. L. C. Comber and John P. Keeves, *Science Education in Nineteen Countries* (New York: Wiley, 1983), p. 299.
2. College Board, "College Board 1984 SAT Scores Rise Four Points, Largest Combined Gain Since 1983," news release, College Board, New York, N.Y., 1984, p. 1.
3. Comber and Keeves, p. 325.
4. Stacey J. Hueftle, Steven J. Rakow, and Wayne W. Welch, *Images of Science* (Minneapolis: Minnesota Research and Evaluation Center, 1983), p. 37. □